Estimating Nonlinear Models with Multiple Fixed Effects: A Computational Note*

LAURA HOSPIDO

Research Division, Bank of Spain, and IZA, Alcalá 48, 28014 Madrid, Spain (e-mail: laura.hospido@bde.es)

Abstract

In this paper we consider estimation of nonlinear panel data models that include multiple individual fixed effects. Estimation of these models is complicated both by the difficulty of estimating models with possibly thousands of coefficients and also by the incidental parameters problem; that is, noisy estimates of the fixed effects when the time dimension is short contaminate the estimates of the common parameters due to the nonlinearity of the problem. We propose a simple variation of existing bias-corrected estimators, which can exploit the additivity of the effects for numerical optimization. We exhibit the performance of the estimators in simulations.

JEL classification numbers: C23, C63.
Keywords: Panel data, nonlinear models, multiple fixed effects, incidental parameters, bias reduction.

I. Introduction

In a typical nonlinear micropanel data model with fixed effects there are hundreds or thousands of individual coefficients to estimate together with a relatively small number of common parameters. A well known computational simplification in the linear model is to obtain first the maximum likelihood (ML) estimates of the common parameters from a regression on the data in deviations from individual means, and secondly retrieve ML estimates of the effects from averaged residuals one by one. A similar computational simplification is available for Newton-Raphson (NR) and related algorithms for nonlinear fixed effects models, which exploits the block-diagonal structure of the Hessian. This simplification has been discussed in Hall (1978), Chamberlain (1980), and Greene (2004) for nonlinear models with a scalar fixed effect. The first purpose of this work is to show how to use an iterated algorithm of this type, the so-called efficient Newton-Raphson iteration (ENR), in a nonlinear model with multiple fixed effects.

*This paper started in collaboration with Manuel Arellano, whose suggestions and encouragement are gratefully acknowledged. I would also like to thank Stéphane Bonhomme, two anonymous referees, the editor of the paper and seminar participants at the SAEe in Granada, the 15th International Conference on Panel Data in Bonn and the EEA-ESEM meetings in Barcelona, for useful suggestions. All errors are mine. The views in this paper are those of the author and do not represent the views of the Bank of Spain or the Eurosystem.

As first noted by Neyman and Scott (1948), when the time series dimension T is small relative to the cross-sectional dimension n, ML estimates of the common parameters can be severely biased, especially in dynamic models. This Incidental Parameters problem arises because the unobserved individual characteristics are replaced by noisy estimates, which bias estimates of model parameters. In particular, the bias of the ML estimator (MLE) is of order $1/T$. In some special cases it is possible to obtain fixed T - large n consistent estimators of certain common parameters, but these situations are more the exception than the rule. Alternatively, a number of additional approaches have been proposed to obtain approximately unbiased estimators as opposed to estimators with no bias at all.[1] One of these approaches consists of estimation from an analytically bias corrected objective function relative to some target criterion.[2] In this paper we also discuss the application of computationally efficient algorithms to modified concentrated likelihoods of this type to obtain estimators without bias to order $1/T$ in nonlinear panel models with multiple fixed effects.

The main contribution of this note is to show how the computational simplification that exploits the block-diagonal structure of the Hessian can be used with bias corrected likelihoods of nonlinear panel data models with multiple fixed effects without affecting the finite sample properties of bias corrected estimators. The estimation of many fixed effects parameters (as many as individuals in the panel) does not pose a real computational problem nowadays for most applications. Computers are now much faster and efficient than at the end of the seventies when the simplification for nonlinear models with a scalar fixed effect was originally discussed. This means that for reduced form panel data models with a fixed effect in the intercept, there is no significant gain in using an iterated algorithm exploiting the block diagonal structure of the Hessian.[3] However, when the model has multiple fixed effects and it has the addition of a modification in the likelihood to correct the incidental parameters problem, the computational simplification matters.[4] And if the model contain any additional complication (like a more structural model) then this simplification will be very helpful.[5]

The paper is organized as follows. Section II introduces the model and notation. Section III explains how the iterated algorithm works. Section IV discusses its application to bias corrected concentrated likelihoods. Section V presents some simulation results. Finally, Section VI concludes. Detailed derivations are given in the Appendix.

[1] See Arellano and Hahn (2007) for a review of this literature on bias-adjusted estimation methods for nonlinear panel data models with fixed effects.

[2] See Pace and Salvan (2006) for adjustments of this type for a generic concentrated likelihood with independent observations, Arellano and Hahn (2007) for static nonlinear panel models and Arellano and Hahn (2006), Bester and Hansen (2009), and Hospido (2010), for the dynamic case. For an automatic way of correcting the bias of the concentrated likelihood see Dhaene and Jochmans (2010).

[3] This can be seen in Section V, in tables 2, 4 and 6 that consider models with one fixed effect.

[4] This is what tables 3, 5 and 7, in Section V, show for models with multiple fixed effects.

[5] We thank one referee for pointing this out.

II. Model and Notation

Let us consider the following model for the joint density of T random vectors conditioned on initial observations, strictly exogenous variables, and fixed effects:

$$f(y_{i1},...,y_{iT} \mid y_{i0},x_{i1},...,x_{iT},\alpha_{i0}) = \prod_{t=1}^{T} f\left(y_{it} \mid y_{i(t-1)},x_{it},\alpha_{i0},\theta_0\right)$$

where θ_0 is a vector of common parameters and α_{i0} is a vector of fixed effects. We observe the random sample $\{y_{i0},...,y_{iT},x_{i0},...,x_{iT}\}_{i=1}^{n}$ and we denote $\alpha_0 = (\alpha'_{10},...,\alpha'_{n0})'$ and $\delta_0 = \left(\theta'_0,\alpha'_0\right)'$. Let the log likelihood of one observation be

$$\ell_{it}(\theta,\alpha_i) = \ln f\left(y_{it} \mid y_{i(t-1)},x_{it},\alpha_i,\theta\right)$$

and let $\ell_i(\theta,\alpha_i) = \sum_{t=1}^{T} \ell_{it}(\theta,\alpha_i)$.

III. Efficient Newton-Raphson iteration

Let us consider the estimator

$$\begin{pmatrix} \widehat{\theta} \\ \widehat{\alpha} \end{pmatrix} = \arg\max_{\theta,\alpha} \sum_{i=1}^{n} \ell_i(\theta,\alpha_i)$$

and let first and second derivatives be denoted by

$$\begin{aligned} d_{\theta i} &= \frac{\partial \ell_i(\theta,\alpha_i)}{\partial\theta}, \qquad d_{\alpha i} = \frac{\partial \ell_i(\theta,\alpha_i)}{\partial\alpha_i} \\ H_{\theta\theta i} &= \frac{\partial^2 \ell_i(\theta,\alpha_i)}{\partial\theta\partial\theta'}, \qquad H_{\alpha\alpha i} = \frac{\partial^2 \ell_i(\theta,\alpha_i)}{\partial\alpha_i\partial\alpha'_i}, \qquad H_{\theta\alpha i} = \frac{\partial^2 \ell_i(\theta,\alpha_i)}{\partial\theta\partial\alpha'_i} \end{aligned}$$

The Kth step of the iteration of a computationally efficient algorithm for obtaining $\widehat{\theta}$ and $\widehat{\alpha}$ takes the form

$$\begin{aligned} \theta_{[K]} - \theta_{[K-1]} &= -\left[\sum_{i=1}^{n}\left(H_{\theta\theta i} - H_{\theta\alpha i}H_{\alpha\alpha i}^{-1}H_{\alpha\theta i}\right)\right]^{-1} \sum_{i=1}^{n}\left(d_{\theta i} - H_{\theta\alpha i}H_{\alpha\alpha i}^{-1}d_{\alpha i}\right) \\ \alpha_{i[K]} - \alpha_{i[K-1]} &= -H_{\alpha\alpha i}^{-1}\left[d_{\alpha i} + H_{\alpha\theta i}\left(\theta_{[K]} - \theta_{[K-1]}\right)\right], \qquad (i = 1,...,n) \end{aligned}$$

where all derivatives are evaluated at $\theta_{[K-1]}$ and $\alpha_{i[K-1]}$.

This result can be easily proved using partitioned inverse formulae (a detailed derivation is in the Appendix). It is a standard result in nonlinear estimation of models with many group effects.

IV. Analytically Adjusted Concentrated Likelihood

When T is short we may be interested to consider an estimator that maximizes a bias corrected concentrated likelihood of the type reviewed in Arellano and Hahn (2007):

$$\widehat{\theta}^{AH} = \arg\max_{\theta} \sum_{i=1}^{n} \left[\ell_i\left(\theta,\widehat{\alpha}_i(\theta)\right) + \beta_i\left(\theta,\widehat{\alpha}_i(\theta)\right)\right]$$

where

$$\widehat{\alpha}_i(\theta) = \arg\max_{\alpha} \ell_i(\theta, \alpha)$$

and $\beta_i(\theta, \alpha_i)$ is an adjustment term.

As long as the adjustment term depends on α, the iterated algorithm discussed above cannot be directly used for estimating $\widehat{\theta}^{AH}$. Note that

$$\begin{pmatrix} \widehat{\theta}^{AH} \\ \widehat{\alpha}^{AH} \end{pmatrix} = \arg\max_{\theta,\alpha} \sum_{i=1}^{n} \left[\ell_i(\theta, \alpha_i) + \beta_i(\theta, \widehat{\alpha}_i(\theta))\right]$$

where $\widehat{\alpha}^{AH} = \widehat{\alpha}\left(\widehat{\theta}^{AH}\right)$. Thus, if we use the analysis of covariance algorithm discussed in the previous section we still need to calculate $\widehat{\alpha}_i(\theta)$ for given values of θ.

A Computationally Effective Estimator

Alternatively, we can consider an estimator of the form

$$\begin{pmatrix} \widetilde{\theta}^{AH} \\ \widetilde{\alpha}^{AH} \end{pmatrix} = \arg\max_{\theta,\alpha} \sum_{i=1}^{n} \left[\ell_i(\theta, \alpha_i) + \beta_i(\theta, \alpha_i)\right]$$

for which the iterated algorithm can be used. This is equivalent to:

$$\widetilde{\theta}^{AH} = \arg\max_{\theta} \sum_{i=1}^{n} \left[\ell_i(\theta, \widetilde{\alpha}_i(\theta)) + \beta_i(\theta, \widetilde{\alpha}_i(\theta))\right] \tag{1}$$

where

$$\widetilde{\alpha}_i(\theta) = \arg\max_{\alpha} \left[\ell_i(\theta, \alpha) + \beta_i(\theta, \alpha_i)\right]$$

The statistic $\widetilde{\alpha}_i(\theta)$ can be regarded as a Bayesian estimator that uses $e^{\beta_i(\theta,\alpha_i)}$ as the prior distribution of α_i for a given value of θ. Thus, under general conditions, $\widetilde{\alpha}_i(\theta)$ will be asymptotically equivalent to $\widehat{\alpha}_i(\theta)$, and $\widetilde{\theta}^{AH}$ will have similar (bias reducing) properties as $\widehat{\theta}^{AH}$ (see Severini, 1998, section 4, for a discussion on the use of adjusted concentrated likelihoods using alternative estimates of nuisance parameters).

It appears that $\widetilde{\theta}^{AH}$ is not only computationally convenient, but it may also exhibit improved finite sample properties in certain situations due to the replacement of $\widehat{\alpha}_i(\theta)$ by $\widetilde{\alpha}_i(\theta)$ (for instance, in regressions of individual effects estimates on strictly exogenous regressors). In fact, correcting the vector of common parameters, θ, and the vector of fixed effects, α_i, both at the same time, was the motivation for the penalty function independently obtained by Bester and Hansen (2009).

Estimation of the Bias

The form of the approximate bias (Arellano and Hahn, 2007) is

$$\beta_i(\theta) \approx \frac{1}{2} tr\left(H_i^{-1}(\theta, \alpha) \Upsilon_i(\theta, \alpha)\right) \tag{2}$$

where

$$H_i(\theta,\alpha) \equiv -\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2 \ell_{it}(\theta,\alpha)}{\partial\alpha\partial\alpha'}$$

$$\Upsilon_i(\theta,\alpha) \equiv \sum_{l=-m}^{m}\omega_{T,l}\Gamma_l(\theta,\alpha)$$

and

$$\Gamma_l(\theta,\alpha) \equiv \frac{1}{T}\sum_{t=\max(1,l+1)}^{\min(T,T+l)}\left[\frac{\partial \ell_{it}(\theta,\alpha)}{\partial\alpha_i}\cdot\frac{\partial \ell_{it-l}(\theta,\alpha)}{\partial\alpha_i'}\right]$$

The quantity m is a bandwidth parameter and $\omega_{T,l}$ denotes a weight that guarantees positive definiteness of $\Upsilon_i(\Gamma,\Theta_i)$.

The bias corrected estimator in (1) that uses an analytical approximation like (2) is equivalent to one of the proposals independently obtained by Bester and Hansen (2009).[6]

Automatically Adjusted Concentrated Likelihood

The half-panel split jackknife provides an automatic way of correcting the bias of the MLE (Dhaene and Jochmans, 2010). The bias corrected estimator is defined as

$$\widehat{\theta}^{DJ} = 2\widehat{\theta} - \overline{\theta}_{1/2}$$

where $\widehat{\theta}$ is the MLE from the full panel, and $\overline{\theta}_{1/2}$ is the average of the two half-panel MLEs, each using $T/2$ time periods and all n cross-sectional units.

V. Monte Carlo Study

The practical importance of the bias corrections depends on how much bias is removed for the small T that is often relevant in econometric applications. However, since the bias-corrected methods used in this paper, either analytically or automatically adjusted, are all asymptotically equivalent, there are no known theoretical reasons to prefer one to another. A particular method may still be preferable for convenience of implementation. In this section, the small-sample performance of the fixed-effects MLE and the bias-corrected estimators is explored in static and dynamic probit models. We present results for different models, keeping the simulation design as consistent as possible across them.[7]

Data Generating Processes

Four probit models are considered:

$$y_{it} = \mathbf{1}[w_{it} + \epsilon_{it} \geq 0]$$

[6] More specifically, to the HS penalty that they consider.

[7] Other studies, that consider nonlinear designs with scalar fixed effects (Carro, 2007; Fernández-Val, 2009), show that the bias in the MLE is similar in magnitude for the logit and the probit models and that bias corrections also perform similarly. Here, we focus on probit designs and extend the analysis to consider multiple fixed effects.

and

$$\epsilon_{it} \sim N(0,1)$$

- Static model with scalar fixed effects:

$$w_{it} = \alpha_{1i0} + \theta_{10} x_{it} + \theta_{20} d_{it}; \quad \left(\theta_0 = [\theta_{10}, \theta_{20}]'; \alpha_{i0} = \alpha_{1i0}\right)$$

- Static model with multiple fixed effects:

$$w_{it} = \alpha_{1i0} + \theta_{10} x_{it} + \alpha_{2i0} d_{it}; \quad \left(\theta_0 = \theta_{10}; \alpha_{i0} = [\alpha_{1i0}, \alpha_{2i0}]'\right)$$

- Dynamic model with scalar fixed effects:

$$w_{it} = \alpha_{1i0} + \theta_{10} x_{it} + \theta_{20} y_{it-1}; \quad \left(\theta_0 = [\theta_{10}, \theta_{20}]'; \alpha_{i0} = \alpha_{1i0}\right)$$

- Dynamic model with multiple fixed effects:

$$w_{it} = \alpha_{1i0} + \theta_{10} x_{it} + \alpha_{2i0} y_{it-1}; \quad \left(\theta_0 = \theta_{10}; \alpha_{i0} = [\alpha_{1i0}, \alpha_{2i0}]'\right)$$

The data were generated with $x_{it} \sim N(0,1)$, $d_{it} = \mathbf{1}[x_{it} + h_{it} > 0]$, and $h_{it} \sim N(0,1)$. For the dynamic designs, the data were generated with $y_{i0} = \mathbf{1}[\alpha_{1i0} + \theta_{10} x_{it} + \epsilon_{i0} > 0]$, and $x_{i0} \sim N(0,1)$, $\epsilon_{i0} \sim N(0,1)$. We set $n = 100$; $T = \{6, 8, 12, 20\}$; $\theta_{10} = 1$, and $\theta_{20} = 0.5$; and ran 1,000 Monte Carlo replications for each design, with just ϵ_{it} redrawn in each replication.

With respect to the individual parameters three alternative scenarios are considered:

(i) the DGP is a model without fixed effects (or, with a little abuse of notation, with constant fixed effects): $\alpha_{1i0} = 0$, $\forall i$; and $\alpha_{2i0} = 0.5$, $\forall i$;

(ii) the DGP is a model with normally distributed fixed effects: $\alpha_{1i0} \sim N(0,1)$; and $\alpha_{2i0} \sim N(0.5, 0.1)$;[8]

(iii) the DGP is a model with fixed effects correlated with both x and d: $\alpha_{1i0} = \frac{1}{\sqrt{3}}(x_{i1} + x_{i2} + x_{i3})$; and $\alpha_{2i0} \sim \left[\frac{0.1}{\sqrt{2}}(\alpha_{1i0} + z_i)\right] + 0.5$, where $z_i \sim N(0,1)$.[9]

Table 1 below summarizes all this information.

[8] The distributional parameters are chosen so that the variability of each term (α_{1i0}, $\alpha_{2i0} d_{it}$, or $\alpha_{2i0} y_{it-1}$) is approximately the same.

[9] The design tries to mimic DGP (ii) in terms of the variability across individuals.

TABLE 1

DGP summary

$y_{it} = \mathbf{1}\left[\alpha_{1i0} + \theta_{10}x_{it} + r_{it} + \epsilon_{it} > 0\right];\ \epsilon_{it} \sim N(0,1)$			
T	$\{6, 8, 12, 20\}$	θ_{10}	1
n	100	α_{1i0} (i):	$0, \forall i$
rep	1,000	α_{1i0} (ii):	$N(0,1)$
x_{it}	$N(0,1)$	α_{1i0} (iii):	$\frac{1}{\sqrt{3}}(x_{i1} + x_{i2} + x_{i3})$
	Static scalar		*Dynamic scalar*
	$r_{it} = \theta_{20}d_{it}$		$r_{it} = \theta_{20}y_{it-1}$
d_{it}	$\mathbf{1}\left[x_{it} + h_{it} > 0\right]$	y_{i0}	$\mathbf{1}\left[\alpha_{1i0} + \theta_{10}x_{i0} + \epsilon_{i0} > 0\right]$
h_{it}	$N(0,1)$	m	1
θ_{20}	0.5	θ_{20}	0.5
	Static multiple		*Dynamic multiple*
	$r_{it} = \alpha_{2i0}d_{it}$		$r_{it} = \alpha_{2i0}y_{it-1}$
d_{it}	$\mathbf{1}\left[x_{it} + h_{it} > 0\right]$	y_{i0}	$\mathbf{1}\left[\alpha_{1i0} + \theta_{10}x_{i0} + \epsilon_{i0} > 0\right]$
h_{it}	$N(0,1)$	m	1
α_{2i0} (i):	$0.5, \forall i$	α_{2i0} (i):	$0.5, \forall i$
α_{2i0} (ii):	$N(0.5, 0.1)$	α_{2i0} (ii):	$N(0.5, 0.1)$
α_{2i0} (iii):	$\left[\frac{0.1}{\sqrt{2}}(\alpha_{1i0} + z_i)\right] + 0.5$	α_{2i0} (iii):	$\left[\frac{0.1}{\sqrt{2}}(\alpha_{1i0} + z_i)\right] + 0.5$
z_i	$N(0,1)$	z_i	$N(0,1)$

Simulation results

We estimate the common parameter θ_0 by maximum likelihood, MLE; applying the analytically bias-corrected estimator of Arellano and Hahn (2006, 2007), AH; and the automatically bias-corrected estimator of Dhaene and Jochmans (2010), DJ; both using the usual Newton-Raphson algorithm (NR) and the efficient version of the iteration (ENR).

Tables 2 to 7 report the effective computation time (in seconds) for each design, along with the median absolute errors and root mean squared errors**.** Failure refers to the percentage of cases of divergence or failure to converge in the nonlinear solution over the 1,000 Monte Carlo replications.[10]

DGP with constant fixed effects

Table 2 reports the results corresponding to the DGP with scalar constant fixed effects and $\theta = (\theta_1, \theta_2)'$.

In the static probit, the MLE of both θ_1 and θ_2 are seriously biased even for $T = 12$. After applying the corrections, the estimates are closer to the true value of the parameters, especially for the AH estimator. In addition, we can see that the ENR algorithm provides a significant computational time improvement with respect to the NR algorithm (from 3.03 to 5.46 times faster). However, when the standard iteration takes 3 seconds there is no really need for a computational simplification.

In the dynamic probit, the MLE of θ_2 is more heavily biased than the one of θ_1. Once again, after applying the corrections, the estimates are closer to the true value of the parameters, but now AH

[10]Those cases are excluded from calculations.

estimator does not always dominate DJ. Also, we can see that the ENR algorithm still provides a substantial improvement in terms of computational time (from 2.99 to 5.32 times faster).

TABLE 2

Probit with scalar constant fixed effects

				NR			*ENR*			*Time*
				MAE	*RMSE*	*Time*	*MAE*	*RMSE*	*Time*	*NR/ENR*
Static	T=6	θ_1	MLE	0.303	0.357	1.633	0.304	0.357	0.336	4.860
			AH	0.168	0.232	2.421	0.167	0.232	0.545	4.442
			DJ	0.196	0.277	0.900	0.178	0.261	0.297	3.030
		θ_2	MLE	0.153	0.251		0.178	0.251		
			AH	0.176	0.200		0.135	0.201		
			DJ	0.134	0.261		0.184	0.270		
	T=12	θ_1	MLE	0.118	0.143	2.342	0.116	0.148	0.444	5.275
			AH	0.052	0.079	3.763	0.051	0.084	0.701	5.368
			DJ	0.089	0.113	1.594	0.082	0.106	0.457	3.488
		θ_2	MLE	0.082	0.150		0.088	0.127		
			AH	0.072	0.085		0.074	0.106		
			DJ	0.082	0.114		0.077	0.112		
	T=20	θ_1	MLE	0.073	0.094	3.307	0.071	0.092	0.620	5.334
			AH	0.040	0.060	5.475	0.039	0.059	1.003	5.459
			DJ	0.044	0.067	2.113	0.040	0.059	0.549	3.849
		θ_2	MLE	0.056	0.084		0.058	0.086		
			AH	0.049	0.075		0.049	0.075		
			DJ	0.061	0.090		0.051	0.076		
Dynamic	T=6	θ_1	MLE	0.269	0.315	1.400	0.269	0.314	0.283	4.947
			AH	0.182	0.238	2.251	0.182	0.238	0.538	4.184
			DJ	0.305	0.482	0.558	0.287	0.459	0.183	3.049
		θ_2	MLE	0.438	0.468		0.440	0.470		
			AH	0.195	0.245		0.198	0.247		
			DJ	0.201	0.299		0.204	0.309		
	T=12	θ_1	MLE	0.114	0.136	2.355	0.114	0.136	0.443	5.316
			AH	0.055	0.083	3.566	0.055	0.082	0.749	4.761
			DJ	0.058	0.086	1.430	0.056	0.083	0.415	3.446
		θ_2	MLE	0.199	0.224		0.202	0.226		
			AH	0.077	0.110		0.080	0.112		
			DJ	0.079	0.119		0.077	0.111		
	T=20	θ_1	MLE	0.065	0.081	3.072	0.065	0.081	0.607	5.061
			AH	0.032	0.049	4.915	0.032	0.049	1.006	4.886
			DJ	0.033	0.049	2.024	0.032	0.049	0.677	2.990
		θ_2	MLE	0.111	0.130		0.116	0.133		
			AH	0.048	0.069		0.050	0.071		
			DJ	0.058	0.085		0.050	0.072		

Notes: MAE=median absolute error, RMSE = root mean squared error, Time = average computation time across replications (in seconds). Scalar fixed effects: α_{1i0} (i) in Table 1.

Table 3 reports the results corresponding to the DGP with multiple constant fixed effects and $\theta = \theta_1$. As expected, with multiple fixed effects the incidental parameter problem gets worse, both for the static and the dynamic probit. Now, the MAE of the MLE is sizable even for values of T such as 12 or 20. Again, the bias-corrected estimators can remove a substantial part of that bias, although the addition of the correction in the likelihood increases the computation time substantially. Interestingly, in this case, the improvements in terms of computational time are very large. These results are encouraging because, in many empirical studies that consider complicated models, the goal is not only to obtain an estimator with a good finite sample performance, but also in a reasonable computing time, especially when bootstrap methods are used for inference.

TABLE 3

Probit with multiple constant fixed effects

				NR			*ENR*			*Time*
				MAE	*RMSE*	*Time*	*MAE*	*RMSE*	*Time*	*NR/ENR*
Static	T=6	θ_1	MLE	0.547	0.618	9.234	0.622	0.683	0.353	26.159
			AH	0.401	0.477	29.207	0.484	0.548	1.156	25.266
			DJ	0.371	27.899	1.599	0.328	0.660	0.072	22.208
		Failure (%)			0.0			0.0		
	T=8	θ_1	MLE	0.407	0.450	12.393	0.425	0.463	0.484	25.605
			AH	0.270	0.319	35.678	0.286	0.330	1.483	24.058
			DJ	0.181	0.309	3.572	0.162	0.251	0.164	21.780
		Failure (%)			0.0			0.0		
	T=12	θ_1	MLE	0.266	0.293	18.118	0.262	0.289	0.736	24.617
			AH	0.158	0.191	48.762	0.154	0.187	2.138	22.807
			DJ	0.114	0.162	6.837	0.082	0.124	0.335	20.409
		Failure (%)			0.0			0.0		
	T=20	θ_1	MLE	0.148	0.163	25.439	0.148	0.164	1.131	22.492
			AH	0.079	0.101	67.290	0.079	0.101	3.219	20.904
			DJ	0.085	0.106	9.896	0.050	0.073	0.636	15.560
		Failure (%)			0.0			0.0		
Dynamic	T=6	θ_1	MLE	0.541	0.604	14.378	0.537	0.601	0.542	26.528
			AH	0.465	0.533	60.186	0.464	0.532	2.342	25.698
			DJ	0.345	21.488	3.383	0.303	51.306	0.160	21.144
		Failure (%)			0.0			1.1		
	T=8	θ_1	MLE	0.400	0.437	14.238	0.398	0.435	0.534	26.663
			AH	0.318	0.355	50.997	0.318	0.355	2.625	19.427
			DJ	0.152	0.241	5.081	0.132	0.216	0.213	23.854
		Failure (%)			0.0			0.3		
	T=12	θ_1	MLE	0.251	0.274	20.291	0.251	0.273	0.743	27.309
			AH	0.164	0.192	62.687	0.164	0.191	2.776	22.582
			DJ	0.150	0.150	8.186	0.073	0.114	0.376	21.771
		Failure (%)			0.0			0.3		
	T=20	θ_1	MLE	0.137	0.149	22.428	0.137	0.148	1.048	21.401
			AH	0.078	0.092	70.917	0.077	0.092	3.322	21.348
			DJ	0.057	0.076	10.012	0.040	0.058	0.656	15.262
		Failure (%)			0.0			0.0		

Notes: MAE=median absolute error, RMSE = root mean squared error, Time = average computation time across replications (in seconds). Multiple fixed effects: α_{1i0} (i) and α_{2i0} (i) in Table 1.

DGP with normal fixed effects

Table 4 reports the results corresponding to the DGP with scalar normal fixed effects and $\theta = (\theta_1, \theta_2)'$.

TABLE 4

Probit with scalar normal fixed effects

				NR			*ENR*			*Time*
				MAE	*RMSE*	*Time*	*MAE*	*RMSE*	*Time*	*NR/ENR*
Static	T=6	θ_1	MLE	0.328	0.388	1.626	0.327	0.388	0.396	4.106
			AH	0.189	0.260	2.304	0.188	0.256	0.655	3.517
			DJ	0.199	0.311	0.857	0.181	0.292	0.300	2.857
		θ_2	MLE	0.212	0.294		0.212	0.295		
			AH	0.154	0.231		0.156	0.232		
			DJ	0.196	0.388		0.200	0.348		
	T=12	θ_1	MLE	0.141	0.173	3.234	0.141	0.172	0.596	5.426
			AH	0.061	0.099	4.705	0.062	0.099	0.935	5.032
			DJ	0.096	0.124	1.907	0.080	0.108	0.507	3.761
		θ_2	MLE	0.097	0.144		0.100	0.145		
			AH	0.081	0.121		0.082	0.122		
			DJ	0.093	0.136		0.088	0.128		
	T=20	θ_1	MLE	0.075	0.099	5.079	0.074	0.098	0.877	5.791
			AH	0.042	0.061	7.852	0.042	0.060	1.392	5.641
			DJ	0.058	0.080	2.871	0.045	0.064	0.736	3.901
		θ_2	MLE	0.066	0.098		0.066	0.099		
			AH	0.058	0.088		0.058	0.088		
			DJ	0.076	0.111		0.063	0.090		
Dynamic	T=6	θ_1	MLE	0.307	0.353	1.257	0.306	0.352	0.291	4.319
			AH	0.231	0.281	2.075	0.231	0.281	0.579	3.584
			DJ	0.209	1.429	0.423	0.202	0.400	0.155	2.729
		θ_2	MLE	0.475	0.516		0.476	0.517		
			AH	0.220	0.284		0.223	0.285		
			DJ	0.209	0.431		0.217	0.390		
	T=12	θ_1	MLE	0.125	0.150	2.916	0.125	0.149	0.545	5.350
			AH	0.062	0.094	4.588	0.062	0.093	0.964	4.759
			DJ	0.077	0.102	1.463	0.072	0.099	0.446	3.280
		θ_2	MLE	0.199	0.224		0.228	0.254		
			AH	0.077	0.110		0.091	0.129		
			DJ	0.079	0.119		0.082	0.123		
	T=20	θ_1	MLE	0.074	0.091	4.788	0.074	0.091	0.798	6.000
			AH	0.037	0.056	7.456	0.037	0.055	1.363	5.470
			DJ	0.036	0.054	2.386	0.034	0.052	0.647	3.688
		θ_2	MLE	0.133	0.158		0.135	0.159		
			AH	0.057	0.087		0.058	0.088		
			DJ	0.080	0.113		0.056	0.084		

Notes: MAE=median absolute error, RMSE = root mean squared error, Time = average computation time across replications (in seconds). Scalar fixed effects: α_{1i0} (ii) in Table 1.

The magnitude of the biases, both in the static and in the dynamic probit, are comparable to those of the previous design in which the fixed effects were constant across individuals. As before, after applying the corrections, the estimates are closer to the true value of the parameters. Again, we can see that the ENR algorithm provides a significant computational time improvement with respect to the NR algorithm (from 2.86 to 5.79 times faster in the static case, and from 2.73 to 6.00 times faster in the dynamic case).

Table 5 reports the results corresponding to the DGP with multiple normal fixed effects and $\theta = \theta_1$. As in Table 3, with multiple fixed effects the incidental parameter problem gets worse, both for the static and the dynamic probit. Again, the MAE of the MLE is sizable even for values of T such as 12 or 20, and the bias-corrected estimators can remove a substantial part of that bias. The improvements in terms of computational time are still very large. Now, however, estimation becomes more unstable, with higher percentages of cases of divergence or failure to converge in the nonlinear solution.

TABLE 5

Probit with multiple normal fixed effects

				NR			*ENR*			*Time*
				MAE	*RMSE*	*Time*	*MAE*	*RMSE*	*Time*	*NR/ENR*
Static	T=6	θ_1	MLE	0.554	0.637	8.174	0.625	0.707	0.330	24.770
			AH	0.412	0.501	26.446	0.488	0.581	1.093	24.196
			DJ	0.436	44.115	1.298	0.386	0.920	0.062	20.935
		Failure (%)			0.2			0.1		
	T=8	θ_1	MLE	0.403	0.455	12.289	0.423	0.470	0.513	23.955
			AH	0.266	0.323	33.744	0.284	0.337	1.613	20.920
			DJ	0.200	0.349	2.977	0.189	0.288	0.155	19.206
		Failure (%)			0.1			0.0		
	T=12	θ_1	MLE	0.281	0.302	19.629	0.281	0.302	0.821	23.909
			AH	0.168	0.196	52.207	0.170	0.197	2.394	21.807
			DJ	0.118	0.172	6.823	0.114	0.172	0.305	22.307
		Failure (%)			0.0			0.0		
	T=20	θ_1	MLE	0.152	0.171	33.358	0.152	0.171	1.359	24.546
			AH	0.077	0.104	88.639	0.078	0.104	3.868	22.916
			DJ	0.094	0.113	12.740	0.089	0.124	0.607	20.988
		Failure (%)			0.0			0.0		
Dynamic	T=6	θ_1	MLE	0.607	0.689	8.478	0.602	0.686	0.308	27.526
			AH	0.540	0.626	40.277	0.539	0.626	1.597	25.220
			DJ	0.353	27.532	2.073	0.343	84.883	0.079	26.240
		Failure (%)			0.0			0.6		
	T=8	θ_1	MLE	0.426	0.477	12.515	0.425	0.475	0.462	27.089
			AH	0.335	0.390	43.959	0.337	0.391	2.491	17.647
			DJ	0.194	0.308	3.715	0.183	0.291	0.139	26.727
		Failure (%)			0.0			0.5		
	T=12	θ_1	MLE	0.276	0.302	20.667	0.275	0.301	0.685	30.171
			AH	0.186	0.217	62.522	0.185	0.217	2.572	24.309
			DJ	0.104	0.155	6.853	0.114	0.170	0.282	24.301
		Failure (%)			0.0			0.4		
	T=20	θ_1	MLE	0.152	0.168	34.830	0.137	0.148	1.138	30.606
			AH	0.086	0.107	103.955	0.077	0.092	3.767	27.596
			DJ	0.072	0.091	10.607	0.040	0.058	0.556	19.077
		Failure (%)			0.0			0.1		

Notes: MAE=median absolute error, RMSE = root mean squared error, Time = average computation time across replications (in seconds). Multiple fixed effects: α_{1i0} (ii) and α_{2i0} (ii) in Table 1.

DGP with correlated fixed effects

Table 6 reports the results corresponding to the DGP with scalar correlated fixed effects and $\theta = (\theta_1, \theta_2)'$.

TABLE 6

Probit with scalar correlated fixed effects

				NR			*ENR*			*Time*
				MAE	*RMSE*	*Time*	*MAE*	*RMSE*	*Time*	*NR/ENR*
Static	T=6	θ_1	MLE	0.373	0.442	1.492	0.370	0.441	0.294	5.075
			AH	0.231	0.311	2.201	0.230	0.310	0.508	4.333
			DJ	0.235	0.397	0.730	0.217	0.376	0.223	3.273
		θ_2	MLE	0.202	0.306		0.203	0.307		
			AH	0.162	0.249		0.166	0.250		
			DJ	0.213	0.427		0.236	0.396		
	T=12	θ_1	MLE	0.149	0.186	2.970	0.147	0.185	0.572	5.192
			AH	0.070	0.108	4.491	0.070	0.107	0.899	4.995
			DJ	0.099	0.135	1.567	0.086	0.119	0.468	3.348
		θ_2	MLE	0.100	0.152		0.101	0.154		
			AH	0.083	0.125		0.083	0.126		
			DJ	0.097	0.146		0.091	0.133		
	T=20	θ_1	MLE	0.088	0.111	5.598	0.088	0.111	0.872	6.420
			AH	0.045	0.068	8.126	0.044	0.068	1.371	5.927
			DJ	0.049	0.071	2.821	0.046	0.065	0.693	4.071
		θ_2	MLE	0.068	0.099		0.069	0.099		
			AH	0.057	0.087		0.058	0.088		
			DJ	0.069	0.102		0.062	0.092		
Dynamic	T=6	θ_1	MLE	0.285	0.353	1.084	0.284	0.353	0.202	5.366
			AH	0.210	0.282	1.679	0.209	0.282	0.394	4.261
			DJ	0.324	11.295	0.393	0.313	0.682	0.121	3.248
		θ_2	MLE	0.498	0.539		0.500	0.540		
			AH	0.241	0.303		0.243	0.305		
			DJ	0.255	36.330		0.253	0.428		
	T=12	θ_1	MLE	0.146	0.175	2.878	0.146	0.174	0.510	5.643
			AH	0.079	0.113	4.557	0.078	0.113	0.900	5.063
			DJ	0.065	0.096	1.247	0.064	0.095	0.380	3.282
		θ_2	MLE	0.229	0.259		0.229	0.259		
			AH	0.093	0.134		0.093	0.135		
			DJ	0.102	0.148		0.089	0.129		
	T=20	θ_1	MLE	0.079	0.098	5.136	0.079	0.098	0.814	6.309
			AH	0.041	0.061	7.948	0.041	0.061	1.373	5.789
			DJ	0.039	0.059	2.404	0.038	0.057	0.623	3.859
		θ_2	MLE	0.133	0.159		0.133	0.159		
			AH	0.059	0.087		0.059	0.087		
			DJ	0.074	0.109		0.057	0.085		

Notes: MAE=median absolute error, RMSE = root mean squared error, Time = average computation time across replications (in seconds). Scalar fixed effects: α_{1i0} (iii) in Table 1.

In this design, in which the fixed effects are correlated with the observed variables, the bias of the MLE is in general bigger than in previous designs. Even in this case, the bias-corrected estimators are able to remove a substantial part of that bias. Once again, the ENR algorithm provides a significant computational time improvement with respect to the NR algorithm (from 3.27 to 6.42 times faster in the static case, and from 3.25 to 6.31 times faster in the dynamic case).

Table 7 reports the results corresponding to the DGP with multiple correlated fixed effects and $\theta = \theta_1$.

TABLE 7

Probit with multiple correlated fixed effects

				NR			ENR			Time
				MAE	*RMSE*	*Time*	*MAE*	*RMSE*	*Time*	*NR/ENR*
Static	T=6	θ_1	MLE	0.609	0.691	7.433	0.603	0.689	0.458	16.229
			AH	0.462	0.551	23.587	0.458	135.614	1.908	12.362
			DJ	0.481	35.400	1.087	0.433	1.033	0.079	13.759
		Failure (%)			0.2			0.2		
	T=8	θ_1	MLE	0.482	0.537	13.458	0.480	0.535	0.590	22.810
			AH	0.343	0.406	38.193	0.346	0.409	1.882	20.294
			DJ	0.217	0.394	3.350	0.237	0.367	0.159	21.069
		Failure (%)			0.0			0.0		
	T=12	θ_1	MLE	0.311	0.342	28.177	0.307	0.337	0.809	34.829
			AH	0.194	0.231	73.644	0.191	0.226	2.384	30.891
			DJ	0.114	0.176	8.970	0.156	0.218	0.283	31.696
		Failure (%)			0.0			0.0		
	T=20	θ_1	MLE	0.159	0.184	40.455	0.160	0.183	1.372	29.486
			AH	0.083	0.114	107.338	0.083	0.114	3.948	27.239
			DJ	0.092	0.116	14.383	0.097	0.135	0.583	24.671
		Failure (%)			0.1			0.0		
Dynamic	T=6	θ_1	MLE	0.515	0.597	6.801	0.511	0.596	0.255	26.671
			AH	0.436	0.526	31.404	0.435	0.530	1.293	24.288
			DJ	0.398	25.833	1.706	0.356	11.191	0.074	23.054
		Failure (%)			0.2			1.1		
	T=8	θ_1	MLE	0.424	0.476	17.030	0.418	0.465	0.509	33.458
			AH	0.333	0.390	66.115	0.330	0.384	2.450	26.986
			DJ	0.192	0.351	5.041	0.176	0.302	0.179	28.162
		Failure (%)			0.0			0.3		
	T=12	θ_1	MLE	0.270	0.304	28.687	0.286	0.311	0.832	34.480
			AH	0.181	0.219	89.077	0.197	0.226	3.145	28.523
			DJ	0.114	0.164	9.097	0.124	0.179	0.344	26.445
		Failure (%)			0.0			0.1		
	T=20	θ_1	MLE	0.150	0.167	45.555	0.152	0.168	1.097	41.527
			AH	0.085	0.108	141.078	0.087	0.108	3.833	36.806
			DJ	0.060	0.087	13.981	0.074	0.104	0.555	25.191
		Failure (%)			0.0			0.0		

Notes: MAE=median absolute error, RMSE = root mean squared error, Time = average computation time across replications (in seconds). Multiple fixed effects: α_{1i0} (iii) and α_{2i0} (iii) in Table 1.

As in the scalar case, also with multiple correlated fixed effects the bias of the MLE is in general bigger than in previous designs. Even though, the bias-corrected estimators are able to remove a comparable part of the bias in this case. Also here, the ENR algorithm provides very large computational time improvement with respect to the NR algorithm.

VI. Conclusions

In this paper we consider estimation of nonlinear panel data models that include multiple individual fixed effects. Estimation of these models is complicated both by the difficulty of estimating models with possibly thousands of coefficients and also by the incidental parameters problem; that is, noisy estimates of the fixed effects when the time dimension is short contaminates the estimates of the common parameters due to the nonlinearity of the problem. We show how to use an iterated algorithm which simplifies estimation in a nonlinear model with multiple fixed effects and we also discuss its application to bias corrected concentrated likelihoods.

Simulations results show that the simplification that exploits the block-diagonal structure of the Hessian not only is computationally convenient, but it also provides adjustments of the likelihood function that result in bias corrected estimators that perform comparably to other bias corrections proposed in the literature. We can think in many microeconometric applications that use nonlinear panel data models. The results of the paper suggest that bias corrected estimates will be very useful in relevant empirical settings given the sample sizes of the panels more often used by researchers and, moreover, because they allow us to introduce more individual heterogeneity to address endogeneity concerns in a robust way.

References

[1] Arellano, M. and Hahn, J. (2006). 'A likelihood-based approximate solution to the incidental parameter problem in dynamic nonlinear models with multiple effects', unpublished manuscript.

[2] Arellano, M. and Hahn, J. (2007). 'Understanding Bias in Nonlinear Panel Models: Some Recent Developments', in R. Blundell, W. Newey, and T. Persson (eds), *Advances in Economics and Econometrics*, Vol. 3, pp. 381-409, Cambridge University Press.

[3] Bester, A. and C. Hansen (2009): "A Penalty Function Approach to Bias Reduction in Nonlinear Panel Models with Fixed Effects", *Journal of Business and Economic Statistics*, 27 (2), 131-148.

[4] Carro, J. (2007). 'Estimating Dynamic Panel Data Discrete Choice Models with Fixed Effects', *Journal of Econometrics*, Vol. 140, pp. 503-528.

[5] Chamberlain, G. (1980). 'Analysis of Covariance with Qualitative Data', *Review of Economic Studies*, Vol. 47, pp. 225–238.

[6] Dhaene, G., and Jochmans, K. (2010). 'Split-panel jackknife estimation of fixed effects models', unpublished manuscript.

[7] Fernández-Val, I. (2009). 'Fixed effects estimation of structural parameters and marginal effects in panel probit models', *Journal of Econometrics,* Vol. 150, pp. 71-85.

[8] Greene, W. (2004). 'The Behaviour of the Maximum Likelihood Estimator of Limited Dependent Variable Models in the Presence of Fixed Effects', *The Econometrics Journal*, Vol. 7, pp. 98–119.

[9] Hall, B. H. (1978). 'A General Framework for Time Series–Cross Section Estimation', *Annales de l'INSEE*, Vol. 30–31, pp. 177–202.

[10] Hospido, L. (2010). 'Modelling Heterogeneity and Dynamics in the Volatility of Individual Wages', *Journal of Applied Econometrics*, forthcoming, DOI: 10.1002/jae.1204.

[11] Neyman, J. and Scott, E. L. (1948). 'Consistent Estimates Based on Partially Consistent Observations', *Econometrica*, Vol. 16, pp. 1-32.

[12] Pace, L. and Salvan, A. (2006). 'Adjustments of the profile likelihood from a new perspective', *Journal of Statistical Planning and Inference*, Vol. 136, pp. 3554-3564.

[13] Severini, T. A. (1998). 'Likelihood Functions for Inference in the Presence of a Nuisance Parameter', *Biometrika,* Vol. 85, pp. 507–522.

Appendix A: Efficient Newton-Raphson iteration

The Kth step of the Newton-Raphson iteration takes the form

$$\delta_K = \delta_{K-1} - \left(\frac{\partial^2 L(\delta_{K-1})}{\partial\delta\partial\delta'}\right)^{-1} \frac{\partial L(\delta_{K-1})}{\partial\delta}$$

or for shortness

$$\Delta\delta = -\left(\frac{\partial^2 L}{\partial\delta\partial\delta'}\right)^{-1} \frac{\partial L}{\partial\delta}$$

where $L(\delta) = \sum_{i=1}^n \ell_i(\theta, \alpha_i)$ and

$$\frac{\partial L}{\partial\delta} = \begin{pmatrix} \frac{\partial L}{\partial\theta} \\ \frac{\partial L}{\partial\alpha_1} \\ \vdots \\ \frac{\partial L}{\partial\alpha_n} \end{pmatrix} = \sum_{t=1}^T \begin{pmatrix} \sum_{i=1}^n \frac{\partial\ell_{it}(\theta,\alpha_i)}{\partial\theta} \\ \frac{\partial\ell_{1t}(\theta,\alpha_1)}{\partial\alpha_1} \\ \vdots \\ \frac{\partial\ell_{nt}(\theta,\alpha_n)}{\partial\alpha_n} \end{pmatrix} = \begin{pmatrix} d_\theta \\ d_\alpha \end{pmatrix}$$

$$\frac{\partial^2 L}{\partial\delta\partial\delta'} = \sum_{t=1}^T \begin{pmatrix} \sum_{i=1}^n \frac{\partial^2\ell_{it}(\theta,\alpha_i)}{\partial\theta\partial\theta'} & \frac{\partial^2\ell_{1t}(\theta,\alpha_1)}{\partial\theta\partial\alpha_1'} & \cdots & \frac{\partial^2\ell_{nt}(\theta,\alpha_n)}{\partial\theta\partial\alpha_n'} \\ \frac{\partial^2\ell_{1t}(\theta,\alpha_1)}{\partial\alpha_1\partial\theta'} & \frac{\partial^2\ell_{1t}(\theta,\alpha_1)}{\partial\alpha_1\partial\alpha_1'} & & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2\ell_{nt}(\theta,\alpha_n)}{\partial\alpha_n\partial\theta'} & 0 & \cdots & \frac{\partial^2\ell_{nt}(\theta,\alpha_n)}{\partial\alpha_n\partial\alpha_n'} \end{pmatrix} = \begin{pmatrix} H_{\theta\theta} & H_{\theta\alpha} \\ H_{\theta\alpha}' & H_{\alpha\alpha} \end{pmatrix}$$

and

$$d_\alpha = \begin{pmatrix} d_{\alpha 1} \\ \vdots \\ d_{\alpha n} \end{pmatrix}, \quad H_{\alpha\alpha} = \begin{pmatrix} H_{\alpha\alpha 1} & & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & H_{\alpha\alpha n} \end{pmatrix}, \quad H_{\theta\alpha} = \begin{pmatrix} H_{\theta\alpha 1} & \dots & H_{\theta\alpha n} \end{pmatrix}$$

so that $d_\theta = \sum_{i=1}^n d_{\theta i}$ and $H_{\theta\theta} = \sum_{i=1}^n H_{\theta\theta i}$, and

$$H_{\theta\alpha} H_{\alpha\alpha}^{-1} = \begin{pmatrix} H_{\theta\alpha 1} H_{\alpha\alpha 1}^{-1} & \dots & H_{\theta\alpha n} H_{\alpha\alpha n}^{-1} \end{pmatrix}$$

$$H_{\theta\alpha} H_{\alpha\alpha}^{-1} H_{\alpha\theta} = \sum_{i=1}^n H_{\theta\alpha i} H_{\alpha\alpha i}^{-1} H_{\alpha\theta i}$$

Letting

$$\begin{pmatrix} H_{\theta\theta} & H_{\theta\alpha} \\ H'_{\theta\alpha} & H_{\alpha\alpha} \end{pmatrix}^{-1} = \begin{pmatrix} H^{\theta\theta} & H^{\theta\alpha} \\ H^{\theta\alpha\prime} & H^{\alpha\alpha} \end{pmatrix}$$

where

$$\begin{aligned} H^{\theta\theta} &= \left(H_{\theta\theta} - H_{\theta\alpha} H_{\alpha\alpha}^{-1} H_{\alpha\theta}\right)^{-1} \\ H^{\theta\alpha} &= -H^{\theta\theta} H_{\theta\alpha} H_{\alpha\alpha}^{-1} \\ H^{\alpha\alpha} &= H_{\alpha\alpha}^{-1} + H_{\alpha\alpha}^{-1} H_{\alpha\theta} H^{\theta\theta} H_{\theta\alpha} H_{\alpha\alpha}^{-1} \end{aligned}$$

the partitioned formula gives

$$\begin{pmatrix} \Delta\theta \\ \Delta\alpha \end{pmatrix} = -\begin{pmatrix} H^{\theta\theta} & H^{\theta\alpha} \\ H^{\theta\alpha\prime} & H^{\alpha\alpha} \end{pmatrix} \begin{pmatrix} d_\theta \\ d_\alpha \end{pmatrix}$$

We have

$$\begin{aligned} H^{\theta\theta} &= \left[\sum_{i=1}^n \left(H_{\theta\theta i} - H_{\theta\alpha i} H_{\alpha\alpha i}^{-1} H_{\alpha\theta i}\right)\right]^{-1} \\ H^{\theta\alpha} d_\alpha &= -H^{\theta\theta} H_{\theta\alpha} H_{\alpha\alpha}^{-1} d_\alpha = -H^{\theta\theta} \sum_{i=1}^n H_{\theta\alpha i} H_{\alpha\alpha i}^{-1} d_{\alpha i} \end{aligned}$$

and

$$-\Delta\theta = H^{\theta\theta} d_\theta + H^{\theta\alpha} d_\alpha = H^{\theta\theta} \left(d_\theta - \sum_{i=1}^n H_{\theta\alpha i} H_{\alpha\alpha i}^{-1} d_{\alpha i}\right)$$

so that

$$\Delta\theta = -\left[\sum_{i=1}^n \left(H_{\theta\theta i} - H_{\theta\alpha i} H_{\alpha\alpha i}^{-1} H_{\alpha\theta i}\right)\right]^{-1} \sum_{i=1}^n \left(d_{\theta i} - H_{\theta\alpha i} H_{\alpha\alpha i}^{-1} d_{\alpha i}\right)$$

Similarly, it is easy to see

$$\Delta\alpha = -H_{\alpha\alpha}^{-1} \left(d_\alpha + H_{\alpha\theta} \Delta\theta\right)$$

so that

$$\Delta\alpha_i = -H_{\alpha\alpha i}^{-1} \left(d_{\alpha i} + H_{\alpha\theta i} \Delta\theta\right), \qquad (i = 1, ..., n)$$